Enumeration and Structure of Trapezoidal Words[☆]

Michelangelo Bucci

Department of Mathematics, University of Turku
FI-20014 Turku, Finland

Alessandro De Luca

Dipartimento di Scienze Fisiche
Università degli Studi di Napoli Federico II
Via Cintia, Monte S. Angelo, I-80126 Napoli, Italy

Gabriele Fici[*]

Laboratoire d'Informatique, Signaux et Systèmes de Sophia-Antipolis
CNRS & Université de Nice-Sophia Antipolis
2000, Route des Lucioles - 06903 Sophia Antipolis cedex, France

Abstract

Trapezoidal words are words having at most $n+1$ distinct factors of length n for every $n \geq 0$. They therefore encompass finite Sturmian words. We give combinatorial characterizations of trapezoidal words and exhibit a formula for their enumeration. We then separate trapezoidal words into two disjoint classes: open and closed. A trapezoidal word is closed if it has a factor that occurs only as a prefix and as a suffix; otherwise it is open. We investigate open and closed trapezoidal words, in relation with their special factors. We prove that Sturmian palindromes are closed trapezoidal words and that a closed trapezoidal word is a Sturmian palindrome if and only if its longest repeated prefix is a palindrome. We also define a new class of words, *semicentral words*, and show that they are characterized by the property that they can be written as $uxyu$, for a central word u and two different letters x, y. Finally, we investigate the prefixes of the Fibonacci word with respect to the property of being open or closed trapezoidal words, and show that the sequence of open and closed prefixes of the Fibonacci word follows the Fibonacci sequence.

Keywords: Sturmian words, trapezoidal words, rich words, closed words, special factors, palindromes, enumerative formula.

1. Introduction

In combinatorics on words, the most famous class of infinite words is certainly that of Sturmian words. They code digital straight lines in the discrete plane having irrational slope, and are

[☆]Part of the results in this paper were presented by the third author at *WORDS 2011, 8th International Conference on Words*, 12-16 September 2011, Prague, Czech Republic [14].

[*]Corresponding author.

Email addresses: michelangelo.bucci@utu.fi (Michelangelo Bucci), alessandro.deluca@unina.it (Alessandro De Luca), gabriele.fici@unice.fr (Gabriele Fici)

Accepted for publication in Theoretical Computer Science *January 22, 2013*

characterized by the property of having exactly $n + 1$ factors of length n, for every $n \geq 0$.

It is well known (see [18], Proposition 2.1.17) that a finite word w is a factor of some Sturmian word if and only if w is a binary *balanced* word, that is, for every letter a and every pair of factors of w of the same length, u and v, one has that u and v contain the same number of a's up to one, i.e.,

$$||u|_a - |v|_a| \leq 1. \tag{1}$$

Finite Sturmian words (finite factors of Sturmian words) have *at most* $n + 1$ factors of length n, for every $n \geq 0$. However, this property does not characterize them, as shown by the word $w = aaabab$, which is not Sturmian since the factors aaa and bab do not verify (1).

The finite words defined by the property of having at most $n + 1$ factors of length n, for every $n \geq 0$, are called *trapezoidal words*. The name comes from the fact that the graph of the function counting the distinct factors of each length (usually called the factor complexity) defines, for trapezoidal words, an isosceles trapezoid.

Trapezoidal words were defined by de Luca [8] by means of combinatorial parameters on the repeated factors of finite words. For a finite word w over a finite alphabet Σ, let K_w be the minimal length of an unrepeated suffix of w and let R_w be the minimal length for which w does not contain right special factors, i.e., factors having occurrences followed by distinct letters. de Luca proved that for any word w of length $|w|$ one always has

$$R_w + K_w \leq |w| \tag{2}$$

and studied the case in which the equality holds. This is in fact the case in which w is trapezoidal, since one can prove (see Proposition 2.8 later) that a word w is trapezoidal if and only if

$$R_w + K_w = |w|. \tag{3}$$

Then, de Luca proved that finite Sturmian words verify (3) and are therefore trapezoidal.

The non-Sturmian trapezoidal words were later characterized by D'Alessandro [6]. Since a non-Sturmian word is not balanced, it must contain a pair of factors having the same length and not verifying (1). Such a pair is called a *pathological pair*. Let w be a non-Sturmian word and (f, g) its pathological pair of minimal length. D'Alessandro proved that w is trapezoidal if and only if $w = pq$ for some $p \in \mathit{Suff}(\{\tilde{z}_f\}^*)$ and $q \in \mathit{Pref}(\{z_g\}^*)$, where $\tilde{z}_f$ is the reversal of the fractional root z_f of f and z_g is the fractional root of g. This characterization is quite involved and will be discussed in detail later.

The characterization by D'Alessandro allows us to give an enumerative formula for trapezoidal words. Indeed, it is known that the number of Sturmian words of length n is equal to

$$S(n) = 1 + \sum_{i=1}^{n} (n - i + 1)\phi(i)$$

(cf. [19, 17]), where ϕ is the Euler totient function, i.e., $\phi(n)$ is the number of positive integers smaller than or equal to n and coprime with n. In Section 3 we prove that the number of non-Sturmian trapezoidal words of length n is

$$T(n) = \sum_{i=0}^{\lfloor (n-4)/2 \rfloor} 2(n - 2i - 3)\phi(i + 2),$$

and thus the number of trapezoidal words of length n is $S(n) + T(n)$ (Theorem 3.2).

We then distinguish trapezoidal words into two distinct classes, accordingly with the definition below.

Definition 1.1. Let w be a finite word over an alphabet Σ. We say that w is *closed* if it is empty or it has a factor (different from the word itself) occurring exactly twice in w, as a prefix and as a suffix, that is, with no internal occurrences. A word which is not closed is called *open*.

For example, the words *aabbaa* and *ababa* are closed, whereas the word *aabbaaa* is open.

The notion of closed word is closely related to the concept of *complete return* to a factor u in a word w, as considered in [15]. A complete return to u in w is any factor of w having exactly two occurrences of u, one as a prefix and one as a suffix. Hence w is closed if and only if it is a complete return to one of its factors; such a factor is clearly both the longest repeated prefix and the longest repeated suffix of w.

The notion of closed word is also equivalent to that of *periodic-like* word [4]. A word w is periodic-like if its longest repeated prefix does not appear in w followed by different letters.

We derive some properties of open and closed trapezoidal words in Section 4. We prove that the set of closed trapezoidal words is contained in that of Sturmian words (Proposition 4.5). We then characterize the special factors of closed trapezoidal words (Lemma 4.6) by showing that every closed trapezoidal word w contains a bispecial factor u such that u is a central word and the left (resp. right) special factors of w are the prefixes (resp. suffixes) of u. We show that trapezoidal palindromes (hence, Sturmian palindromes) are closed words (Theorem 4.9), and that a closed trapezoidal word is a Sturmian palindrome if and only if its longest repeated prefix is a palindrome (Proposition 4.12).

In Section 5 we introduce *semicentral words*, that are trapezoidal words in which the longest repeated prefix, the longest repeated suffix, the longest left special factor and the longest right special factor all coincide. We prove in Theorem 5.2 that a trapezoidal word w is semicentral if and only if it is of the form $w = uxyu$, for a central word u and two different letters x, y.

We then study, in Section 6, the prefixes of the Fibonacci infinite word f, and show that the sequence of the numbers of consecutive prefixes of f that are closed (resp. open) is the sequence of Fibonacci numbers (Theorem 6.1).

The paper ends with conclusions and open problems in Section 7.

2. Trapezoidal words

An *alphabet*, denoted by Σ, is a finite set of symbols, called *letters*. A *word* over Σ is a finite sequence of letters from Σ. The subset of the alphabet Σ constituted by the letters appearing in w is denoted by *alph*(w). We let w_i denote the i-th letter of w. Given a word $w = w_1 w_2 \cdots w_n$, with $w_i \in \Sigma$ for $1 \leq i \leq n$, the nonnegative integer n is the *length* of w, denoted by $|w|$. The empty word has length zero and is denoted by ε. The set of all words over Σ is denoted by Σ^*. The set of all words over Σ having length n is denoted by Σ^n. For a letter $a \in \Sigma$, $|w|_a$ denotes the number of a's appearing in w.

The word $\tilde{w} = w_n w_{n-1} \cdots w_1$ is called the *reversal* of w. A *palindrome* is a word w such that $\tilde{w} = w$. In particular, we assume $\tilde{\varepsilon} = \varepsilon$ and so the empty word is a palindrome.

A word z is a *factor* of a word w if $w = uzv$ for some $u, v \in \Sigma^*$. In the special case $u = \varepsilon$ (resp. $v = \varepsilon$), we call z a prefix (resp. a suffix) of w. An occurrence of a factor of a word w is

internal if it is not a prefix nor a suffix of w. The set of prefixes, suffixes and factors of the word w are denoted respectively by *Pref*(w), *Suff*(w) and *Fact*(w). A *border* of the word w is any word in *Pref*$(w) \cap$ *Suff*(w) different from w.

The *factor complexity* of a word w is the function defined by $f_w(n) = |Fact(w) \cap \Sigma^n|$, for every $n \geq 0$.

A factor u of w is *left special* (resp. *right special*) in w if there exist $a, b \in \Sigma$, $a \neq b$, such that $au, bu \in Fact(w)$ (resp. $ua, ub \in Fact(w)$). A factor is *bispecial* in w if it is both left and right special.

A *period* for the word w is a positive integer p, with $0 < p \leq |w|$, such that $w_i = w_{i+p}$ for every $i = 1, \ldots, |w| - p$. Since $|w|$ is always a period for w, we have that every non-empty word has at least one period. We can unambiguously define *the* period of the word w as the smallest of its periods. The period of w is denoted π_w. The *fractional root* z_w of a word w is its prefix whose length is equal to the period of w.

A word w is *a power* if there exists a non-empty word u and an integer $n > 1$ such that $w = u^n$. A word which is not a power is called *primitive*.

For example, let $w = aababba$. The left special factors of w are ε, a, ab, b and ba. The right special factors of w are ε, a, ab and b. Therefore, the bispecial factors of w are ε, a, ab and b. The period of w is $\pi_w = 6$ and the fractional root of w is $z_w = aababb$.

The following parameters were introduced by Aldo de Luca [8]:

Definition 2.1. Given a word $w \in \Sigma^*$, H_w (resp. K_w) denotes the minimal length of a prefix (resp. a suffix) of w which occurs only once in w.

Definition 2.2. Given a word $w \in \Sigma^*$, L_w (resp. R_w) denote the minimal length for which there are no left special factors (resp. no right special factors) of that length in w.

Example 2.3. Let $w = aaababa$. The longest left special factor of w is aba, and it is also the longest repeated suffix of w; the longest right special factor of w is aa, and it is also the longest repeated prefix of w. Thus, we have $L_w = 4$, $K_w = 4$, $R_w = 3$ and $H_w = 3$.

Notice that the values H_w, K_w, L_w and R_w are positive integers, unless for the words that are powers of a single letter, in which case $R_w = L_w = 0$ and $K_w = H_w = |w|$. Moreover, one has $f_w(R_w) = f_w(L_w)$ and $\max\{R_w, K_w\} = \max\{L_w, H_w\}$ (cf. [8], Corollary 4.1).

The following proposition is from de Luca ([8], Proposition 4.2).

Proposition 2.4. *Let w be a word of length $|w|$ such that $|alph(w)| > 1$ and set $m_w = \min\{R_w, K_w\}$ and $M_w = \max\{R_w, K_w\}$. The factor complexity f_w is strictly increasing in the interval $[0, m_w]$, is nondecreasing in the interval $[m_w, M_w]$ and strictly decreasing in the interval $[M_w, |w|]$. Moreover, for i in the interval $[M_w, |w|]$, one has $f_w(i + 1) = f_w(i) - 1$. If $R_w < K_w$, then f_w is constant in the interval $[m_w, M_w]$.*

Proposition 2.4 allows one to give the following definition.

Definition 2.5. A non-empty word w is *trapezoidal* if

- $f_w(i) = i + 1 \quad \text{for} \quad 0 \leq i \leq m_w$,
- $f_w(i + 1) = f_w(i) \quad \text{for} \quad m_w \leq i \leq M_w - 1$,

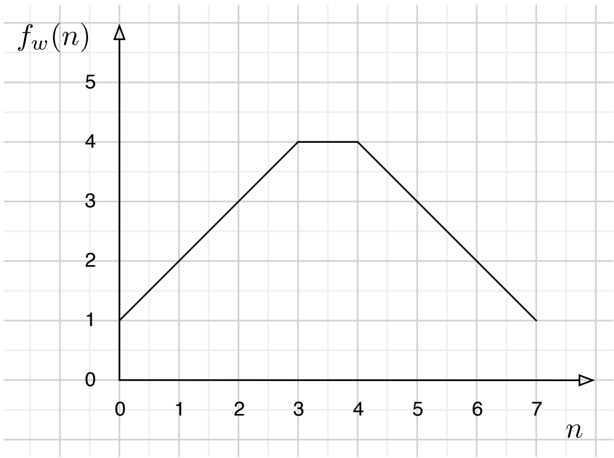

Figure 1: The graph of the complexity function of the trapezoidal word $w = aaababa$. One has $\min\{R_w, K_w\} = 3$ and $\max\{R_w, K_w\} = 4$.

- $f_w(i+1) = f_w(i) - 1 \quad \text{for} \quad M_w \leq i \leq |w|$.

Trapezoidal words were considered for the first time by de Luca [8]. The name *trapezoidal* was given by D'Alessandro [6]. The choice of the name is motivated by the fact that for these words the graph of the complexity function defines an isosceles trapezoid (possibly degenerated in a triangle).

Remark 2.6. It follows from the definition that a trapezoidal word is necessarily a binary word, since $f_w(1)$, the number of distinct letters appearing in w, must be 1 or 2.

Example 2.7. The word $w = aaababa$ considered in Example 2.3 is trapezoidal. See Figure 1.

In the following proposition we gather some characterizations of trapezoidal words.

Proposition 2.8. *Let w be a word. The following conditions are equivalent:*

1. *w is trapezoidal;*
2. *$|w| = L_w + H_w$;*
3. *$|w| = R_w + K_w$;*
4. *w is binary and has at most one left special factor of length n for every $n \geq 0$;*
5. *w is binary and has at most one right special factor of length n for every $n \geq 0$;*
6. *w has at most $n+1$ distinct factors of length n for every $n \geq 0$;*
7. *$|f_w(n+1) - f_w(n)| \leq 1$ for every $n \geq 0$.*

Proof. The equivalence of (1), (2), (3), (4) and (5) is in [8] and [6]. The equivalence of (5) and (6) follows from elementary considerations on the factor complexity of binary words. Indeed, it is easy to see the number of distinct factors of length $n+1$ of a binary word w is at most equal to the number of distinct factors of length n plus the number of right special factors of length n. The equivalence of (7) and (1) comes directly from the definitions and from Proposition 2.4. □

From now on, we fix the alphabet $\Sigma = \{a, b\}$.

Recall that a word $w \in \Sigma^*$ is *Sturmian* if and only if it is balanced, i.e., verifies (1). The following proposition is from de Luca ([8], Proposition 7.1).

Proposition 2.9. *Every Sturmian word is trapezoidal.*

The inclusion in Proposition 2.9 is strict, since there exist trapezoidal words that are not Sturmian, e.g. the word $w = aaababa$ considered in Example 2.3.

Recall that a word w is *rich* [15] (or *full* [1]) if it has $|w| + 1$ distinct palindromic factors, that is the maximum number of distinct palindromic factors a word can contain. The following proposition is from de Luca, Glen and Zamboni ([11], Proposition 2).

Proposition 2.10. *Every trapezoidal word is rich.*

Again, the inclusion in Proposition 2.10 is strict, since there exist binary rich words that are not trapezoidal, e.g. the word $w = aabbaa$.

We now discuss the case of non-Sturmian trapezoidal words. We first recall the definition and some properties of *central words* [12].

Definition 2.11. A word $w \in \Sigma^*$ is a central word if and only if it has two coprime periods p and q and length equal to $p + q - 2$.

Proposition 2.12 ([16, 3])**.** *A word $w \in \Sigma^*$ is a central word if and only if it satisfies one of the following equations, for some $u, v \in \Sigma^*$ and $n \geq 0$:*

1. $w = a^n$,
2. $w = b^n$,
3. $w = uabv = vbau$.

Moreover, if case 3 holds one has that u and v are central words.

It is known that if $w = uabv = vbau$ is a central word, then the longest between u and v has not internal occurrences in w. Thus, w is a closed word. Note that this is also a consequence of our Theorem 4.9.

Proposition 2.13 ([7])**.** *A word $w \in \Sigma^*$ is a central word if and only if w is a palindrome and awa and bwb are both Sturmian words.*

Proposition 2.14 ([12])**.** *A word $w \in \Sigma^*$ is a central word if and only if there exists a Sturmian word w' such that w is a bispecial factor of w'.*

Recall that a word w is not balanced if and only if it contains a *pathological pair* (f, g), that is, f and g are factors of w of the same length but they do not verify (1). Moreover, if f and g are chosen of minimal length, there exists a palindrome u such that $f = aua$ and $g = bub$, for two different letters a and b (see [18, Proposition 2.1.3] or [5, Lemma 3.06]).

We can also state that f and g are Sturmian words, since otherwise they would contain a pathological pair shorter than $|f| = |g|$ and hence w would contain such a pathological pair, against the minimality of f and g. So the word u is a palindrome such that aua and bub are Sturmian words, that is, by Proposition 2.13, u is a central word.

Moreover, we have the following

Lemma 2.15. *Let w be a binary non-Sturmian word and (f, g) a pathological pair of w of minimal length. Then there exists a unique central word u such that $f = aua$ and $g = bub$.*

Proof. Suppose by contradiction that there exists a word $u' \neq u$, with $|u| = |u'|$, such that w contains as factors aua, bub, $au'a$ and $bu'b$. Let h be the longest common prefix between u and u'. Since $u \neq u'$ there exist distinct letters $x, y \in \Sigma$ such that hx is a prefix of u and hy is a prefix of u'. Then w would contain as factors aha, ahb, bha and bhb. Therefore, w would contain a pathological pair (aha, bhb) shorter than (f, g), a contradiction. □

The following lemma is attributed to Aldo de Luca in [6]. We provide a proof for the sake of completeness.

Lemma 2.16. *Let w be a non-Sturmian word and (f, g) the pathological pair of w of minimal length. Then f and g do not overlap in w.*

Proof. By contradiction, suppose that w contains a factor uvz such that $f = uv$ and $g = vz$, with $|v| > 0$. Since $||f|_a - |g|_a| = ||u|_a - |z|_a|$, then (u, z) would be a pathological pair of w shorter than (f, g). □

We then have:

Proposition 2.17. *Let w be a binary non-Sturmian word. Then there exists a unique central word of minimal length u such that $w = v_1auav_2bubv_3$, for $a, b \in \Sigma$ distinct letters and $v_1, v_2, v_3 \in \Sigma^*$. We call the word u the* central root *of w.*

The following is the characterization of trapezoidal non-Sturmian words given by D'Alessandro [6].

Theorem 2.18. *Let w be a binary non-Sturmian word. Then w is trapezoidal if and only if*

$$w = pq, \qquad \text{with } p \in \mathit{Suff}(\{\tilde{z}_f\}^*), \quad q \in \mathit{Pref}(\{z_g\}^*) \tag{4}$$

where (f, g) is the pathological pair of w of minimal length. In particular, if w is trapezoidal, then $K_w = |q|$ and the longest right special factor of w is the prefix of w of length $R_w - 1$, that is, the longest proper prefix of p.

Hence, trapezoidal words are either Sturmian or of the form described in Theorem 2.18.

Example 2.19. Let $w = aaababa$ be the non-Sturmian trapezoidal word considered in Example 2.3. We have $f = aaa$ and $g = bab$, so that $\tilde{z}_f = a$ and $z_g = ba$. The word w can be written as $w = pq$, with $p = aaa$ and $q = baba$.

Remark 2.20. From now on, according to Theorem 2.18, (f, g) is the pair of pathological factor of w of minimal length.

We now show that the words p and q of the factorization (4) are Sturmian words. We premise the following known result:

Proposition 2.21 (cf. [10])**.** *Let w be a finite word and z_w its fractional root. Then w is Sturmian if and only if so is every power of z_w.*

Lemma 2.22. *Let w be a trapezoidal non-Sturmian word. Then the words p and q of the factorization (4) are Sturmian words.*

Proof. Since f and g are Sturmian, by the preceding proposition all powers of z_f and z_g are Sturmian as well. Observe that $p \in \mathit{Suff}(\{\tilde{z}_f\}^*)$ can be restated as $\tilde{p} \in \mathit{Pref}(\{z_f\}^*)$. As prefixes and reversals of Sturmian words are Sturmian, the assertion follows. □

The following result, due to de Luca, Glen and Zamboni [11], states that trapezoidal palindromes are all Sturmian. We give a new proof based on Theorem 2.18.

Theorem 2.23. *The following conditions are equivalent:*

1. *w is a trapezoidal palindrome;*
2. *w is a Sturmian palindrome.*

Proof. Let w be a trapezoidal palindrome. If w is non-Sturmian we can write, by Theorem 2.18, $w = pq = v_1 f g v_2 = \tilde{v}_2 g f \tilde{v}_1$, with $v_1, v_2 \in \Sigma^*$ such that $p = v_1 f$ and $q = g v_2$. Without loss of generality, we can assume $|v_1| \geq |v_2|$; then $p = v_1 f$ contains f and g as factors, a contradiction since, by Lemma 2.22, p is Sturmian. So w cannot be a non-Sturmian word.

Hence, we proved that trapezoidal palindromes are Sturmian. Since by Proposition 2.9, Sturmian words are trapezoidal, the assertion follows. □

3. An enumerative formula for trapezoidal words

Mignosi [19] proved the following formula (see also [17]) for the number $S(n)$ of Sturmian words of length n:

$$S(n) = 1 + \sum_{i=1}^{n} (n - i + 1)\phi(i), \tag{5}$$

where ϕ is the Euler totient function.

Therefore, by Proposition 2.9, in order to count the number of trapezoidal words of length n, it is sufficient to count the number of non-Sturmian trapezoidal words of that length. Recall from Theorem 2.18 that every non-Sturmian trapezoidal word w can be uniquely written as $w = pq$, with $p \in \mathit{Suff}(\{\tilde{z}_f\}^*)$ and $q \in \mathit{Pref}(\{z_g\}^*)$, with $f = aua$, $g = bub$, where u is the central root of w and (f, g) is the pathological pair of w of minimal length. Our goal is to count, for a given central word u, how many non-Sturmian trapezoidal words exist with central root u and length n.

Let us premise the following technical lemma, which is a straightforward consequence of Theorem 2.18.

Lemma 3.1. *Let u be a central word, and $x, y \in \Sigma$ different letters. The non-Sturmian trapezoidal words having central root u are exactly the words of the form pq, for any $p \in \mathit{Suff}(\{\tilde{z}_{xux}\}^*)$ with $|p| \geq |u| + 2$, and $q \in \mathit{Pref}(\{z_{yuy}\}^*)$ with $|q| \geq |u| + 2$.*

Theorem 3.2. *For all $n \geq 0$, the number of trapezoidal words of length n is*

$$1 + \sum_{i=1}^{n} (n - i + 1)\phi(i) \; + \sum_{i=0}^{\lfloor (n-4)/2 \rfloor} 2(n - 2i - 3)\phi(i + 2).$$

Proof. We first notice that for any word z and any $j \geq 0$, there exists a unique word of length j in $\mathit{Suff}(\{\tilde{z}\}^*)$ (resp. in $\mathit{Pref}(\{z\}^*)$). From this remark and Lemma 3.1, we deduce that for any $n \geq 4$ and for any central word u of length $0 \leq |u| \leq \lfloor (n-4)/2 \rfloor$, there are exactly $2(n - 2(|u| + 2) + 1) =$

$2(n-2|u|-3)$ non-Sturmian trapezoidal words of length n having central root u. Since there are $\phi(i+2)$ central words of length i (see [12]), we have that the number of non-Sturmian trapezoidal words of length n is

$$T(n) = \sum_{i=0}^{\lfloor (n-4)/2 \rfloor} 2(n-2i-3)\phi(i+2). \tag{6}$$

Since the number of trapezoidal words of length n is $S(n)+T(n)$, by (5) and (6) the statement follows. □

In Table 1 we give the number of trapezoidal words of each length up to 20.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$Trap \cap \Sigma^n$	2	4	8	16	28	46	70	102	140	190	250	318	398	496	602	724	862	1018	1192	1382

Table 1: The number of trapezoidal words of each length up to 20.

4. Open and closed trapezoidal words

In this section we derive some properties of open and closed trapezoidal words. Recall that a word w is closed if it is empty or it has a factor (different from w) occurring exactly twice in w, as a prefix and as a suffix; otherwise it is open.

Remark 4.1. Let w be a non-empty word over Σ. The following characterizations of closed words follow easily from the definition:

1. w has a factor occurring exactly twice in w, as a prefix and as a suffix of w;
2. the longest repeated prefix of w does not have internal occurrences in w, that is, occurs in w only as a prefix and as a suffix;
3. the longest repeated suffix of w does not have internal occurrences in w, that is, occurs in w only as a suffix and as a prefix;
4. the longest repeated prefix of w is not right special in w;
5. the longest repeated suffix of w is not left special in w;
6. w has a border that does not have internal occurrences in w;
7. the longest border of w does not have internal occurrences in w;
8. w is the complete return to its longest repeated prefix;
9. w is the complete return to its longest border;
10. $w = uv = zu$, with v, z non-empty, and $Fact(w) \cap \Sigma u \Sigma = \emptyset$.

Proposition 4.2. *Let w be a word. Then w is open (resp. closed) if and only if $\tilde{w}$ is open (resp. closed).*

Proof. Straightforward. □

The following proposition gives a characterization of open trapezoidal words.

Proposition 4.3. *Let w be a trapezoidal word. Then the following conditions are equivalent:*

1. *w is open;*
2. *the longest repeated prefix of w is also the longest right special factor of w;*
3. *the longest repeated suffix of w is also the longest left special factor of w.*

Proof. (1) $\Rightarrow$ (2). Let h_w be the longest repeated prefix of w and x the letter such that $h_w x$ is a prefix of w. Since w is not closed, h_w has a second non-suffix occurrence in w followed by a letter y. Since h_w is the longest repeated prefix of w, we have $y \neq x$. Therefore, h_w is right special in w.

Suppose that w has a right special factor r longer than h_w. Since w is trapezoidal, w has at most one right special factor for each length (Proposition 2.8). Since the suffixes of a right special factor are right special factors, we have that h_w must be a proper suffix of r. Since r is right special in w, it has at least two occurrences in w followed by different letters. This implies a non-prefix occurrence of $h_w x$ in w, against the definition of h_w.

(2) $\Rightarrow$ (1). Let h_w be the longest repeated prefix of w. Since h_w is a right special factor of w there exist different letters a and b such that $h_w a$ and $h_w b$ are factors of w. This implies that h_w cannot appear in w only as a prefix and as a suffix, that is, w cannot be a closed word.

(1) $\Leftrightarrow$ (3) can be proven symmetrically. □

Proposition 4.4. *Let w be a trapezoidal word. If w is open, then $H_w = R_w$ and $K_w = L_w$. If w is closed, then $H_w = K_w$ and $L_w = R_w$.*

Proof. The claim for open trapezoidal words follows directly from Proposition 4.3.

Suppose that w is a closed trapezoidal word. Then $H_w = K_w$ (since w is closed) and therefore $L_w = R_w$ (since w is trapezoidal). □

The latter proposition, however, does not characterize open and closed trapezoidal words. Actually, one can have $H_w = K_w = L_w = R_w$, as is the case in the word *abba*, which is closed, and also in the word *aaba*, which is open.

Open trapezoidal words can be Sturmian (e.g. *aaabaa*) or not (e.g. *aaabab*). Closed trapezoidal words, instead, are always Sturmian, as shown in the following proposition, that can also be found in a paper of the first two authors together with Aldo de Luca (cf. [2], Proposition 3.6). Here we report a proof following a totally different approach, based on Theorem 2.18.

Proposition 4.5. *Let w be a trapezoidal word. If w is closed, then w is Sturmian.*

Proof. Suppose that w is not Sturmian. Then, by Theorem 2.18, $w = pq$, $p \in Suff(\{\tilde{z}_f\}^*)$, $q \in Pref(\{z_g\}^*)$, with (f, g) being the pathological pair of w of minimal length, $K_w = |q|$ (and hence $R_w = |p|$ since w is trapezoidal) and the longest right special factor of w is the prefix of w of length $R_w - 1$.

Let k_w be the suffix of q of length $K_w - 1$, that is, the longest repeated suffix of w. Since w is closed, the only other occurrence of k_w in w is as a prefix of w.

If $R_w \geq K_w$, then k_w is a prefix of the longest right special factor of w. This is a contradiction with the fact that k_w does not have internal occurrences in w.

If $R_w < K_w$, then p is a prefix of k and hence p is a factor of q. This implies that f is a factor of q, and therefore q contains both f and g as factors. Hence q would be non-Sturmian, contradicting Lemma 2.22. □

As a corollary of Proposition 4.5, we have that every trapezoidal word is open or Sturmian. We now focus on closed trapezoidal words and their special factors.

Lemma 4.6. *Let w be a closed trapezoidal word and u the longest left special factor of w. Then u is also the longest right special factor of w (and thus u is a bispecial factor of w). Moreover, u is a central word.*

Proof. Let u be the longest left special factor of w, so that there exist different letters $a, b \in \Sigma$ such that au and bu are factors of w.

We claim that au and bu both occur in w followed by some letter. Indeed, suppose the contrary. Then one between au and bu, say au, appears in w only as a suffix. Let k_w be the longest repeated suffix of w. Since u is a repeated suffix of w, we have $|k_w| \geq |u|$. If $|k_w| = |u|$, then $k_w = u$ and since w is closed, u appears in w only as a prefix and as a suffix, against the hypothesis that bu is a factor of w. So $|k_w| > |u|$ and therefore au must be a suffix of k_w. This implies an internal occurrence of au in w, a contradiction.

So, there exist letters x, y such that aux and buy are factors of w. Now, we must have $x \neq y$, since otherwise ux would be a left special factor of w longer than u. Thus u is right special in w. Since w is closed, we have, by Proposition 4.4, $L_w = R_w$, and therefore u is the longest right special factor of w.

By Proposition 4.5, w is Sturmian. Since u is a bispecial factor of a Sturmian word, we have by Proposition 2.14 that u is a central word. □

Example 4.7. Let $w = aababaaba$. The longest repeated prefix of w is $aaba$, which is also the longest repeated suffix and does not have internal occurrences, hence w is a closed trapezoidal word.

The longest left special factor of w is aba, which is also the longest right special factor of w and it is a central word.

By Theorem 2.23, trapezoidal palindromes coincide with Sturmian palindromes, so the next results will be stated in terms of Sturmian palindromes but one can replace the word "Sturmian" with the word "trapezoidal".

Some results on Sturmian palindromes can be found in [9]. In particular, we recall here the following one.

Theorem 4.8 ([9], Theorem 29)**.** *A palindrome $w \in \Sigma^*$ is Sturmian if and only if its minimal period $\pi(w)$ satisfies $\pi_w = R_w + 1$.*

The next theorem shows that Sturmian palindromes, and so, by Theorem 2.23, trapezoidal palindromes, are all closed words.

Theorem 4.9. *Let w be a Sturmian palindrome. Then w is closed.*

Proof. By contradiction, suppose that w is open. Then h_w, the longest repeated prefix of w, is also the longest right special factor of w (Proposition 4.3). Since w is a palindrome, we have that the longest repeated suffix of w is $\tilde{h}_w$, the reversal of h_w. In particular, then, $K_w = H_w$.

By Proposition 4.4, $H_w = R_w$ and $K_w = L_w$. Thus we have $H_w = R_w = K_w = L_w = |w|/2$, since w is trapezoidal (see Proposition 2.8). It follows that $w = h_w xx\tilde{h}_w$, for a letter $x \in \Sigma$. By Theorem 4.8, the period of w is $R_w + 1 = |h_w xx|$, so that $\tilde{h}_w = h_w$. Therefore, we have $w = h_w xx h_w$.

Since h_w is right special in w, there exists a letter $y \neq x$ such that $h_w y$ is a factor of w. Now, any occurrence of $h_w y$ in w cannot be preceded by the letter x, since $h_w = \tilde{h}_w$ is the longest repeated suffix of w. Thus w contains the factor $y h_w y$.

Hence, w contains both $h_w xx$ and $y h_w y$ as factors, and this contradicts the fact that w is Sturmian. □

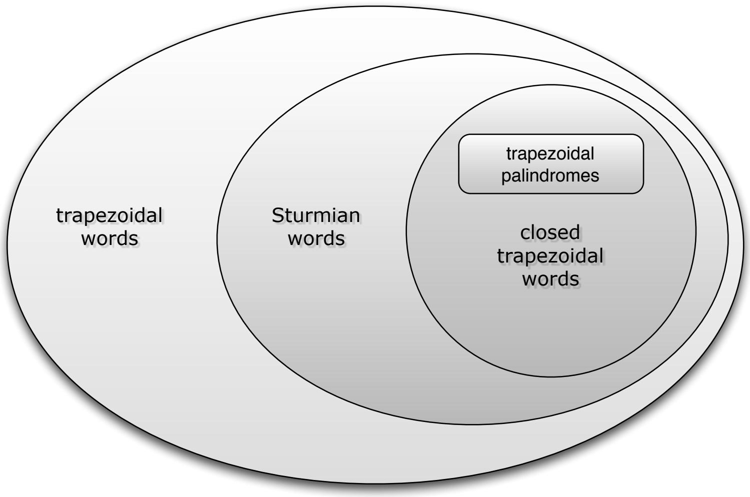

Figure 2: A Venn diagram illustrating the inclusion relations of trapezoidal words.

From Theorem 4.9 and Lemma 4.6, we derive the following characterization of the special factors of Sturmian palindromes.

Corollary 4.10. *Let w be a Sturmian palindrome. Then the longest left special factor of w is also the longest right special factor of w and it is a central word.*

Let us recall the following characterization of rich words from [15]:

Proposition 4.11. *A word w is rich if and only if every factor of w which is a complete return to a palindrome is palindromic itself.*

We have the following result:

Proposition 4.12. *Let w be a closed trapezoidal word and h_w its longest repeated prefix. Then w is a Sturmian palindrome if and only if h_w is a palindrome.*

Proof. The word w is a complete return to h_w, and it is rich by Proposition 2.10. Since trapezoidal palindromes coincide with Sturmian palindromes by Theorem 2.23, the preceding proposition implies the assertion. □

The general structure of a closed trapezoidal word is depicted in Figure 3.

Remark 4.13. Using Proposition 4.11, one can prove that Theorem 4.9 can be generalized to rich palindromes. Indeed, let w be a rich palindrome, and u the longest border of w. In order to prove

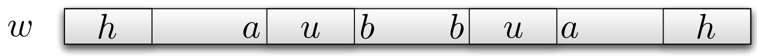

Figure 3: The structure of a closed trapezoidal word. The longest repeated prefix, h_w, is also the longest repeated suffix and does not have internal occurrences. The longest left special factor, u, is also the the longest right special factor and it is a central word (Lemma 4.6). The word is a palindrome if and only if h_w is a palindrome (Proposition 4.12).

that w is closed, it is sufficient to prove that u does not have internal occurrences in w. Suppose the contrary. Then w would contain a proper prefix of the form uzu. By Proposition 4.11, uzu is a palindrome. Hence, uzu would be a suffix of w, against the maximality of u.

Notice that the converse is not true. The word $w = aaababbaabbabaaa$ is a closed palindrome of length 16 but contains only 16 palindromic factors, so it is not rich.

5. Semicentral words

In what follows, we let by h_w, k_w, r_w, l_w respectively denote the longest repeated prefix, the longest repeated suffix, the longest right special factor and the longest left special factor of the word w. An intriguing class of trapezoidal words is that of words in which all these factors coincide.

Definition 5.1. Let w be a trapezoidal word. We say that w is *semicentral* if $h_w = k_w = r_w = l_w$.

Notice that, by Proposition 4.3, semicentral words can also be defined as those open trapezoidal words in which the longest repeated prefix is also the longest repeated suffix.

The following theorem provides a characterization of semicentral words.

Theorem 5.2. *A trapezoidal word w is semicentral if and only if there exists a central word u such that $w = uxyu$ with $x, y \in \Sigma$, $x \neq y$.*

Proof. By (3), if w is semicentral we can write $w = uxyu$, with $u = h_w = k_w = r_w = l_w$ and x, y (possibly equal) letters in Σ. Since u is a bispecial factor of w, it must have a third occurrence in w; as $u = h_w = k_w$, such an occurrence cannot be preceded by y or followed by x. We have then

$$w = uxyu = \alpha\bar{y}u\bar{x}\beta \tag{7}$$

where $\Sigma = \{x, \bar{x}\} = \{y, \bar{y}\}$, $\alpha, \beta \in \Sigma^*$ and $|\alpha| + |\beta| = |u|$. If $\alpha = u$, then $uxyu = u\bar{y}u\bar{x}$, so that $x \neq y$ and $yu = uy$. Thus $u = y^n$ for some $n \geq 0$, so that it is a central word. Similarly, if $\beta = u$ it follows that $x \neq y$ and $u = x^n$ is central.

Let then $0 < |\alpha|, |\beta| < |u|$. From (7) it follows that there exists a prefix u_1x of u such that $ux = \alpha\bar{y}u_1x$, and a suffix yu_2 of u such that $yu = yu_2\bar{x}\beta$. Hence, as $|u| = |\alpha| + |\beta|$, we obtain

$$u = u_1xyu_2 = u_2\bar{x}\bar{y}u_1 \ .$$

Again, this implies $x \neq y$, otherwise the number of occurrences of x in u_1xyu_2 and $u_2\bar{x}\bar{y}u_1$ would be different. Thus, $u = u_1xyu_2 = u_2yxu_1$ is central by Proposition 2.12.

Conversely, let u be a central word and $w = uxyu$, with $\Sigma = \{x, y\}$. We claim that xuy occurs in w. This is clear if u is a power of x or y; by Proposition 2.12, the other possibility is $u = u_1xyu_2 = u_2yxu_1$ for some words u_1, u_2, so that

$$w = uxyu = u_2y \cdot xu_1xyu_2y \cdot xu_1$$

gives the desired occurrence of xuy. Hence u is a right special factor and a repeated suffix of w, so that $R_w \geq |u| + 1$ and $K_w \geq |u| + 1$. Since $|w| = 2(|u| + 1) \leq R_w + K_w$, it follows from (2) that w is trapezoidal and $k_w = r_w = u$. Analogously, since u is a left special factor and a repeated prefix of w, one proves that $h_w = l_w = u$. □

By Proposition 2.12, every central word u different from a power of a single letter can be written as $u = u_1xyu_2 = u_2yxu_1$ for central words u_1 and u_2 and distinct letters x, y. Theorem 5.2 provides a similar characterization for semicentral words (and this motivates the choice of the name).

It is also worth noticing that semicentral words are non-strictly bispecial Sturmian words (see [13]), that is, for any semicentral word w one has that aw, bw, wa and wb are all Sturmian words, but nevertheless either awb or bwa is not Sturmian.

Example 5.3. The semicentral words of length 8 are: *aaaabaaa*, *aaabaaaa*, *abaababa*, *ababaaba*, *bababbab*, *babbabab*, *bbbabbbb* and *bbbbabbb*.

As a consequence of Theorem 5.2, we have that the number $SC(n)$ of semicentral words of length n is:

$$SC(n) = \begin{cases} 0, & \text{if } n \text{ is odd,} \\ 2\phi\left(\frac{n}{2} + 1\right), & \text{if } n \text{ is even.} \end{cases}$$

6. Open and closed prefixes of the Fibonacci word

In this section, we give an interesting example of the open/closed dichotomy in trapezoidal words, by analyzing prefixes of the Fibonacci infinite word under this new light. The Fibonacci infinite word f is the word

$$f = abaababaabaababaababaabaaba\cdots$$

obtained as the limit of the substitution $a \mapsto ab$, $b \mapsto a$. Note that every prefix of the Fibonacci word is Sturmian, and therefore trapezoidal.

We investigate which prefixes of f are open and which are closed. If one writes up the list of the prefixes of f for each length (see Table 6), can notice that the sequences of "open" and "closed" alternate following the Fibonacci sequence F_i (defined by $F_1 = F_2 = 1$ and $F_{i+2} = F_{i+1} + F_i$ for every $i \geq 1$). More precisely, the numbers of consecutive prefixes that are closed (resp. open), that is, the lengths of the runs of "c" and "o" in Table 6, are the Fibonacci numbers. In Theorem 6.1 we show this fact.

n	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
$f_{1\cdots n}$	c	o	c	o	c	c	o	o	c	c	c	o	o	o	c	c	c	c	c	o

Table 2: Which prefixes of the Fibonacci word are closed and which are open.

Theorem 6.1. *Let w be a prefix of the Fibonacci word f. Then w is open if and only if there exists i such that $F_{i+1} - 1 \leq |w| \leq 2F_i - 2$.*

Proof. First, let us recall some well known properties of the Fibonacci word. Let $S = \{\varepsilon, a, aba, abaaba, abaababaa$ be the set of palindromic prefixes of f. It is well known that the elements of S are the prefixes s_i of f such that $|s_i| = F_i - 2$, $i \geq 3$. The words in S are all central words. Also, for every $i \geq 4$ we have

$$s_{i+1} = s_i x_i y_i s_{i-1} = s_{i-1} y_i x_i s_i, \tag{8}$$

where $x_i = b$, $y_i = a$ for even i, and $x_i = a$, $y_i = b$ for odd i.

Our assertion is trivially verified if $|w| \leq 3 = F_5 - 2$. Hence we only need to show that for all $i \geq 4$, a prefix w of f with $F_{i+1} - 1 \leq |w| < F_{i+2} - 1$ is

1. open if $F_{i+1} - 1 \leq |w| \leq 2F_i - 2$, and
2. closed otherwise, i.e., when $2F_i - 1 \leq |w| \leq F_{i+2} - 2$.

Let us fix $i \geq 4$ and set $x = x_i$, $y = y_i$ as in (8). We know that the prefix of f of length $F_{i+1} - 2$ is s_{i+1}. By Theorem 4.9, it is a closed word and its longest repeated prefix is s_i. Moreover, s_i is also the longest repeated prefix of the semicentral word $s_i xy s_i$, that is the prefix of f of length $2F_i - 2$. Therefore, s_i is the longest repeated prefix for all w's in between, i.e., for each $w \in \mathit{Pref}(f)$ with $F_{i+1} - 1 \leq |w| \leq 2F_i - 2$. As s_i has an internal occurrence in all such prefixes, they are all open, proving point 1.

It remains to show that any prefix of

$$s_{i+2} = s_i xy s_{i+1} = s_i xy s_i xy s_{i-1}$$

longer than $s_i xy s_i$ is closed. Indeed, such words have a border which is strictly longer than s_i, and which cannot have any internal occurrences for otherwise $s_i x$ would have a second occurrence in $s_i xy s_i$. $\square$

prefix of f	length	open/closed	example
$s_i xy s_{i-1} = s_{i+1}$	$F_{i+1} - 2$	closed	*abaababaaba*
$s_i xy s_{i-1} y$	$F_{i+1} - 1$	open	*abaababaabaa*
$s_i xy s_{i-1} yx$	F_{i+1}	open	*abaababaabaab*
...	...	...	...
$s_i xy s_i$	$2F_i - 2$	open	*abaababaabaaba*
$s_i xy s_i x$	$2F_i - 1$	closed	*abaababaabaabab*
$s_i xy s_i xy$	$2F_i$	closed	*abaababaabaababa*
...	...	...	...
$s_i xy s_{i+1} = s_{i+2}$	$F_{i+2} - 2$	closed	*abaababaabaababaaba*
$s_i xy s_{i+1} x$	$F_{i+2} - 1$	open	*abaababaabaababaabab*

Table 3: The structure of the prefixes of f with respect to the palindromic prefixes s_i.

Corollary 6.2. *The longest word in a run of closed prefixes of f is a central word. The longest word in a run of open prefixes of f is a semicentral word.*

7. Conclusions and open problems

In this paper we have investigated trapezoidal words, which are a natural generalization of finite Sturmian words. We exhibited a formula for counting trapezoidal words and studied open and closed trapezoidal words separately.

Trapezoidal words form a subclass of the class of rich words, that are words containing the maximum number of palindromic factors. A challenging problem is that of finding an enumerative formula for rich words. This problem is still open, even in the binary case. We think that separating rich words in open and closed could give further insights on these words.

More generally, we think that the open/closed dichotomy can be useful in the study of other classes of words. For instance, it would be interesting to generalize the results obtained in Section 6 for the Fibonacci word to the class of standard words.

References

[1] S. Brlek, S. Hamel, M. Nivat, C. Reutenauer, On the palindromic complexity of infinite words, International Journal of Foundations of Computer Science 15 (2004) 293–306.

[2] M. Bucci, A. de Luca, A. De Luca, Rich and Periodic-Like Words, in: V. Diekert, D. Nowotka (Eds.), DLT 2009, 13th International Conference on Developments in Language Theory, Vol. 5583 of Lecture Notes in Comput. Sci., Springer, 2009, pp. 145–155.

[3] A. Carpi, A. de Luca, Central Sturmian Words: Recent Developments, in: C. De Felice, A. Restivo (Eds.), DLT 2005, 9th International Conference on Developments in Language Theory, Vol. 3572 of Lecture Notes in Comput. Sci., Springer, 2005, pp. 36–56.

[4] A. Carpi, A. de Luca, Periodic-like words, periodicity and boxes, Acta Informatica 37 (2001) 597–618.

[5] E. M. Coven, G. A. Hedlund, Sequences with Minimal Block Growth, Mathematical Systems Theory 7 (2): 138–153 (1973).

[6] F. D'Alessandro, A combinatorial problem on Trapezoidal words, Theoret. Comput. Sci. 273 (1-2) (2002) 11–33.

[7] A. de Luca, Sturmian words: structure, combinatorics, and their arithmetics, Theoret. Comput. Sci. 183 (1997) 45–82.

[8] A. de Luca, On the combinatorics of finite words, Theoret. Comput. Sci. 218 (1999) 13–39.

[9] A. de Luca, A. De Luca, Combinatorial properties of Sturmian palindromes, Internat. J. Found. Comput. Sci. 17 (3) (2006) 557–573.

[10] A. de Luca, A. De Luca, Some characterizations of finite Sturmian words, Theoret. Comput. Sci. 356 (2006) 118–125.

[11] A. de Luca, A. Glen, L. Q. Zamboni, Rich, Sturmian, and trapezoidal words, Theoret. Comput. Sci. 407 (2008) 569–573.

[12] A. de Luca, F. Mignosi, Some combinatorial properties of Sturmian words, Theoret. Comput. Sci. 136 (2) (1994) 361–385.

[13] G. Fici. A Characterization of Bispecial Sturmian Words, in: B. Rovan, V. Sassone, and P. Widmayer (Eds.), MFCS 2012, 37th International Symposium on Mathematical Foundations of Computer Science, Vol. 7464 of Lecture Notes in Comput. Sci., Springer, 2012, pp. 383–394.

[14] G. Fici, A Classification of Trapezoidal Words, in: P. Ambrož, Š. Holub, Z. Masáková (Eds.), WORDS 2011, 8th International Conference on Words, no. 63 in Electronic Proceedings in Theoretical Computer Science, 2011, pp. 129–137.

[15] A. Glen, J. Justin, S. Widmer, L. Q. Zamboni, Palindromic richness, European J. Combin. 30 (2009) 510–531.

[16] L. Ilie, W. Plandowski, Two-variable word equations, RAIRO Theor. Inform. Appl. 34 (2000) 467–501.

[17] E. P. Lipatov, A classification of binary collections and properties of homogeneity classes. (Russian), Problemy Kibernet. 39 (1982) 67–84.

[18] M. Lothaire, Algebraic Combinatorics on Words, Encyclopedia of Mathematics and its Applications, Cambridge Univ. Press, New York, NY, USA, 2002.

[19] F. Mignosi, On the number of factors of Sturmian words, Theoret. Comput. Sci. 82 (1991) 71–84.